SPEECH RECOGNITION DEVICE

BACKGROUND OF THE INVENTION

The present invention relates to a speech recognition device. More particularly, the present invention relates to semiconductor integrated circuits to perform speech recognition.

In recognizing speech and images, clustering and labeling are basic processes, and self-organizing clustering has been proposed in reference document 1 and a clustering system employing a learning method with a teacher has been proposed in reference document 2 and reference document 3. The reference documents are described below. Speech recognition using this system has also been reported. Although parallel processing digital LSIs to perform the self-organizing clustering process at a high speed have been proposed, a problem, that the area of chips is increased in a parallel processing system, occurs. As analog circuits that can calculate distance, and can be realized with a small number of devices, a circuit that uses neuron MOSFETs and calculates a Manhattan distance has been proposed in reference document 4, and that which puts out the square of an Euclidean distance has been proposed in reference document 5.

Reference document 1 is, Y. Miyanaga, S. Okumura, and K. Tochinai, [On versatility and adaptability of self-organizing clustering] Electronic Information/Communication Conference (A), vol. J75-A, no. 7, pp. 1207-1215, July 1992.

Reference document 2 is, Y. Miyanaga and K. Tochinai, [On high speed and high accurate learning of network by self-organization and teacher] Electronic Information/Communication Conference (A), vol. J78-A, no. 11, pp.1475-1484, Nov. 1995.

Reference document 3 is, R. Islam, Y. Miyanaga, and

K. Tochinai, [Multi-clustering network for data classification system] IEICE Trans. Fundamentals, vol. E80-A, no. 9, pp. 1647-1654, Sep. 1997.

Reference document 4 is, M. Konda, T. Shibata, and T. Ohmi, [Neuron-MOS correlator based on Manhattan distance computation for event recognition hardware] IEEE International Symposium on Circuit and Systems, vol. 4, Atlanta, USA, pp. 217-220, May 1996.

Reference document 5 is, U. Cilingiroglu and D. Y. Aksin, [A 4-transistor Euclidean distance cell for analog classifiers] IEEE International Symposium on Circuits and Systems, vol. 1, California, USA, pp. 84-87, May 1998.

The present applicants have examined the parallel operation processing digital LSI using the above-mentioned speech recognition art, but have been confronted with a problem in that the number of basic operation modules becomes very large and the chip area of integrated circuit becomes large. Therefore, while aiming at reduction in circuit scale, the applicants have tried to realize clustering and labeling, which are basic processes in the above-mentioned speech and image recognition, in analog circuits.

SUMMARY OF THE INVENTION

The object of the present invention is to provide a speech recognition device that can realize speech recognition using a small-scale circuit. The other object of the present invention is to provide a speech recognition device appropriate to semiconductor integrated circuits. These objects and their new characteristics will be made clear by the description of the present specification and accompanying drawings.

Typical constitutions among those to be disclosed in the present invention are briefly explained below. Similarity circuits, which receive input signals composed of multi-dimensional vectors corresponding to the spectrum envelope of speech inputs to be recognized and put out characteristics based on the self-organizing

algorithm, calculate a distance for a dimension using a pair of neuron MOSFETs corresponding to each dimension in order to obtain distances between the above-mentioned multi-dimensional input vectors and pattern vectors prepared in advance for speech recognition, perform the clustering process by summing the currents that flow through each neuron MOSFET and forming a voltage signal that corresponds to the degree of similarity, supply the voltage signal to a matrix circuit for matrix operation in which capacitors corresponding to weighting operations are arranged in matrix, and perform the labeling process by outputting what is most similar to the patterns, prepared in advance among the matrix operation outputs, as the recognition result.

BRIEF DESCRIPTION OF THE DRAWINGS

The features and advantages of the invention will be more clearly understood from the following description taken in conjunction with the accompanying drawings, in which:

FIG.1 is a general structure diagram that shows an embodiment of the speech recognition device relating to the present invention.

FIG.2 is a general signal processing flow chart that shows an embodiment in the speech recognition device relating to the present invention.

FIG.3 is a general circuit diagram that shows an embodiment of the speech recognition device (clustering/labeling circuit) relating to the present invention.

FIG.4 is a circuit diagram that shows an embodiment of the similarity circuit used in the present invention.

FIG.5 is a diagram that illustrates the operation principles of the neuron MOSFET used in the present invention.

FIGs.6A and 6B are circuit diagrams that illustrate how to operate the neuron MOSFET used in the present invention.

FIG.7 is a circuit diagram that shows an embodiment of the operational amplifier circuit used in the present invention.

FIG.8 is a circuit diagram that shows an embodiment of the C-matrix used in the present invention.

FIGs.9A and 9B are circuit diagrams that illustrate how to operate the C-matrix circuit shown in FIG.8.

FIG.10 is a table that shows an embodiment of the capacitance values (fF) of the template values C1ij of the clustering layer when the five vowels are recognized by the speech recognition device relating to the present invention.

FIG.11 is a table that shows an embodiment of the learning results of weight and the capacitance values (fF) of the C-matrix of the labeling layer when the five vowels are recognized by the speech recognition relating to the present invention.

FIG.12 is a diagram that shows the waveforms of the simulation results when the five vowels are supplied to the speech recognition device relating to the present invention.

FIG.13 is a diagram that shows the output waveforms of the simulation results when the five vowels are supplied to the speech recognition device relating to the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The general structure diagram of an embodiment of the speech recognition device relating to the present invention is shown in FIG.1. The speech recognition system in this embodiment comprises two layers. The first layer, that is the clustering layer, puts out characteristics based on the self-organizing algorithm according to the input vector y consisting of p dimensions. The second layer, that is the labeling layer, receives the characteristic outputs formed in the first clustering layer, to which weights based on the teacher-attached algorithm are multiplied and summed. By the way,

in the above-mentioned reference document 2, recognition and learning is carried out simultaneously in the same system as that shown in FIG.1, but it is difficult to perform this in analog circuits.

In this embodiment, therefore, coefficients calculated in advance by a computer are embedded in a chip and the chip is made to only perform recognition using these values. Expressions used for recognition are shown below. There are m cluster nodes in the first layer and each node has a pattern vector xi (i = 1, 2, ..., m). Each node calculates the similarity Si (i = 1, 2, ..., m) based on the Euclidean distance Di (i = 1, 2, ..., m) between the p-dimensional input vector y = (y1, y2, ..., yp) and the pattern vector xi = (xi 1, xi2, ..., xip) as follows.

$$D_i = \sqrt{\sum_{j=1}^{p} (y_j - x_{ij})^2}, \qquad \ldots (1)$$

$$S_i = \begin{cases} 1 - (D_i / D_s)^2 & D_i < D_s \\ 0 & D_i \geq D_s \end{cases} \qquad \ldots (2)$$

In expression 2, Ds is a threshold provided to deal with non-linear problems.

The second layer has n nodes and output Si of the first layer is multiplied by m-dimensional weight vector wt = (wt1, wt2, ..., wtm) (t = 1, 2, ..., n) and summed. Output z = (z1, z2, ..., zn) of the system is the sign.

$$R_t = \sum_{i=1}^{m} w_{ti} S_i \qquad \ldots (3)$$

$$z_t = \begin{cases} 0 & R_i < 0 \\ 1 & R_t \geq 0 \end{cases} \qquad \ldots (4)$$

The learning of the network is determined by configuring a software system that performs the identical operations and using the method described in the above-mentioned reference document 2. Although not restricted particularly in this embodiment, the component of xi is rounded to a whole number between 1 and 255 for hardware

use and an appropriately rounded whole number is used for wt because of the limitations by the chip design rule.

The general signal processing flow chart in an embodiment of the speech recognition device relating to the present invention is shown in FIG.2. Although not restricted particularly in this embodiment, circuits to recognize the five vowels a, i, u, e, and o are used for example in the following description.

The recognized speech input signal forms a signal consisting of multi-dimensional vector that corresponds to the spectrum envelope by the envelope processing after obtaining the frequency spectrum of the speech signal pitched in four levels using, for example, the linear predictive analysis method (ARMA speech analysis method), although not restricted particularly. From thus formed input signals, the speech recognition signals label:/a/, /i/, /u/, /e/, and /o/ are formed in the clustering/labeling circuit, which will be described below.

The general circuit diagram of an embodiment of the speech recognition device (clustering/labeling circuit) relating to the present invention is shown in FIG.3. In the structure of this embodiment, m p-dimensional similarity circuits are arranged in parallel and n x m C (capacitor) matrix is attached to the outputs of these similarity circuits. In this figure, black boxes x11 - xmp that constitute the similarity circuits are composed of pairs of neuron MOSFETs in the distance circuits. The components of the similarity circuit inputs are connected to each other, and input voltages are supplied to all the distance circuits simultaneously. The pattern vector Xi is memorized in each similarity circuit as a ratio of the capacitances and the result of the similarity operation is supplied to the C-matrix, then the weighting operation and sign discrimination are performed.

As described above, when the five vowels (a, i, u, e, o) are recognized, the black boxes x11 to xmp that

constitute the similarity circuits in the embodiment are composed of 30 x 16 units. In other words, the input signals Vin1 to Vinp are set to the input signals Vin1 to Vin30 consisting of the 30-dimensional vectors that correspond to the spectrum envelope, and supplied to the pairs of neuron MOSFETs shown by the 16-black boxes in which the input signals Vin1 to Vin30 are arranged in the direction of column. By this, the output signals Vs1 to Vsm formed in the clustering layer are set to 16 signals such as Vs1 to Vs16.

In the C-matrix circuit, the 16 rows that correspond to the 16 output signals from the above-mentioned similarity circuits, the six columns, that is, the five columns that correspond to the five vowels (a, i, u, e, o) and the comparison capacitor column, and the dummy capacitor Cdum to equalize the total capacitance in each column, are provided. Therefore, in total, in the C-matrix, 17 x 6 capacitors are provided.

In this embodiment, the neuron MOSFETs are used for a subtractive operation to calculate distances in the similarity circuits (clustering circuit), as mentioned above. The diagram that illustrates the operation principles of the neuron MOSFET is shown in FIG.5. To the gate of the neuron MOSFET, n inputs of capacitors are connected. According to the operation principles of the neuron MOSFET, Vi (i=1, 2, ..., n) is applied to each input first and then the switch is closed to pre-charge 0V to the gate. Next, the switch is opened to terminate the pre-charge and the input voltage is changed to Vi' (i=1, 2, ..., n). The voltage applied to the gate of MOSFET at this time is as shown in expression 5.

$$V_{gs} = \frac{\sum_{i=1}^{n} C_i(V_i' - V_i)}{C_{all}} \qquad \ldots (5)$$

"Call" is the total capacitance of the capacitors attached to the gate.

The basic characteristic of the MOSFET used in the

circuit in the embodiment is as follows. In the range Vthn < Vgsn < Vdsn + Vthn, the n-channel MOSFET operates in the saturation area and the relation between the drain current and the gate current is as shown in expression 6.

$$I_{dsn} = \frac{KP_n}{2}(V_{gsn} - V_{thn})^2 \quad \ldots (6)$$

The p-channel MOSFET operates in the linear area (non-saturation area) in the range Vdsp + Vthp > Vgsp, as shown in expression 7.

$$I_{dsp} = -KP_p\left\{(V_{gsp} - V_{thp})V_{dsp} - \frac{1}{2}V_{dsp}^2\right\} \quad \ldots (7)$$

In expressions 6 and 7, Vgsn, Vdsn, Vthn, KPn, and Idsn refer to the gate - source voltage, the drain - source voltage, the threshold voltage, the transconductance, and the drain current, respectively, of the n-channel MOSFET, and Vgsp, Vdsp, Vthp, KPp, and Idsp refer to the gate - source voltage, the drain - source voltage, the threshold voltage, the transconductance, and the drain current, respectively, of the p-channel MOSFET. In the present embodiment, the degree of similarity is calculated by combining the saturation area of the n-channel MOSFET and the linear area of the p-channel MOSFET, as described later.

The circuit diagram of the similarity circuit in the embodiment used in the present invention is shown in FIG.4. In the circuit of the present embodiment, a circuit, which calculates the distance between the p-dimensional input vector y = (y1, y2, ..., yp) and the pattern vector xi = (xi1, xi2, ..., xip), is schematically shown as a typical one. As described above, when the five vowels are recognized, similar circuits, 16 in total, are provided.

Although not restricted in particular, it is assumed that the components of the above-mentioned vectors y and xi are whole numbers between 0 and 255. In the present embodiment, the two neuron MOSFTEs calculate the value

corresponding to one dimension. Each of the j-th pair of neuron MOSFETS has capacitance of C1ij, C2ij, and C3. C1ij and C2ij are determined using the j-th component xij of the pattern vector xi so as to have the ratio as shown in the following expression.

$$C_{1ij} : C_{2ij} = x_{ij} : 255 - x_{ij} \quad \ldots (8)$$

C3 is set as shown in expression 9, being made to correspond to the threshold voltage of the n-channel MOSFET.

$$C_3 = C_{all} \frac{V_{thn}}{V_{dd}} \quad \ldots (9)$$

"Call" is the total sum of capacitances of the capacitors attached to the gate, as in expression 5.

As the input voltage, the analog voltage Vinj for each element of the vector is given by expression 10.

$$V_{inj} = \frac{y_j}{255} V_{dd} \quad \ldots (10)$$

The voltage of the node is kept equal to the voltage Vbias of the reversed input of the operational amplifier circuit, because all the outputs (drains) of the neuron MOSFET pairs are connected to each other and the node is provided with feedback from the operational amplifier circuit through the p-channel MOSFET. In other words, the operational amplifier circuit forms the output voltage so that the voltage Vbias given to the reversed input (-) becomes equal to the non-reversed input (+) voltage, that is, the voltage of the drain of the neuron MOSFET is equal to that of the drain of the p-channel MOSFET at the connection node and, then, it drives the p-channel MOSFET. It is possible, thereby, to establish the operation conditions with which the neuron MOSFET is driven in the saturation area and the p-channel MOSFET is driven in the linear area.

The circuit diagrams the illustrate the operation method of the neuron MOSFET are shown in FIG.6A and FIG.6B. FIG.6A shows the pre-charge cycle, during which the n-channel MOSFET attached to the floating gate is

turned on and a pre-charge is performed to the grounding voltage 0V of the circuit. During the pre-charge cycle, the capacitors C1ij and C2ij of the neuron MOSFET on the left-hand side are provided with the input voltage Vinij, and the capacitor C3, with 0 V. On the contrary, the capacitor C1ij of the neuron MOSFET on the right-hand side is provided with Vdd and C2ij and C3, with 0 V.

FIG.6B shows the execute cycle, during which the n-channel MOSFET attached to the above-mentioned floating gate is turned off and the capacitor C3 is provided with Vdd. During the execute cycle, in contrast with the above-mentioned case, the capacitors C1ij and C2ij of the neuron MOSFET on the right-hand side are provided with the input voltage Vinij. On the contrary, the capacitor C1ij of the neuron MOSFET on the left-hand side is provided with Vdd, and C2ij, with 0 V. At this time, The voltage Vgsn (left) and Vgsn (right) between the gate and source of the left- and right-hand side neuron MOSFETs in the cell are obtained as expressions 11 and 12 by substituting expressions 8, 9, and 10 into expression 5.

$$V_{gsn(left)} = V_{thn} - \frac{C_0}{C_{all}} \frac{(y_j - x_{ij})}{255} V_{dd}, \qquad \ldots (11)$$

$$V_{gsn(right)} = V_{thn} + \frac{C_0}{C_{all}} \frac{(y_j - x_{ij})}{255} V_{dd} \qquad \ldots (12)$$

Since either Vgsn (left) or Vgsn (right) in the above-mentioned expressions is smaller than Vthn, the drain current does not flow in such a case because of the cut off state. The drain current flows in the other MOSFET and if the gate voltage is smaller than Vbias + Vthn, expression 13 is obtained from expression 6

$$I_{dsn} = \frac{KP_n}{2} \left\{ \frac{C_0}{C_{all}} \frac{(y_j - x_{ij})}{255} V_{dd} \right\}^2 \qquad \ldots (13)$$

When the gate voltage exceeds Vbias + Vthn, expression 13 does not hold because the neuron MOSFET operates in the linear area. In the simulation that will be shown later, however, it does not matter even if the

squared current cannot be obtained because the area moves to the area beyond the threshold Ds in expression 2.

Switching of the input signal Vinij as shown in FIG.6A and FIG.6B is performed by the switch circuit SW in FIG.3. The capacitor C3 and the n-channel switch MOSFET are provided with the same operation signal. Therefore, in the circuit in FIG.3, the circuit to control the capacitor C3 and the n-channel switch MOSFET is omitted.

In FIG.4, since no current flows through the input of the operational amplifier circuit, all the drain current of the neuron MOSFET flows into the p-channel MOSFET. The current that flows in the p-channel MOSFET is the sum of the current of all the neuron MOSFETs in the same row, therefore, expression 14 is obtained.

$$-I_{dsp} = \sum_{j=1}^{p} \frac{KP_n}{2}\left\{\frac{C_0}{C_{all}}\frac{(y_j - x_{ij})}{255}V_{dd}\right\}^2 + I_0 \quad \ldots (14)$$

Here, the constant current I_0, which is provided to the drain of the p-channel MOSFET, also serves to keep the feedback by conducting current to the p-channel MOSFET during the pre-charge cycle. On the other hand, feedback is applied to the p-channel MOSFET via the operational amplifier circuit, a gate voltage corresponding to the drain current that flows is applied to with the aid of the operational amplifier circuit, and the gate voltage is used as output.

The circuit diagram of an embodiment of the above-mentioned operational amplifier circuit is shown in FIG.7. The drains of the n-channel differential MOSFETs M5 and M7 are provided with a load circuit composed of the p-channel MOSFETs M4 and M5, which are arranged in the current mirror layout, and the source commonly connected to the above-mentioned MOSFETs M5 and M7 is provided with the n-channel current source MOSFET M8 that conducts the operation current. The output signal obtained from the drain of the above-mentioned

differential MOSFET M7 is sent to the gate of the p-channel amplification MOSFET M11. The drain of the amplification MOSFET M11 is provided with the n-channel current source MOSFET M12 as a load.

The drain output of the amplification MOSFET M11 is commonly supplied to the gates of the n-channel source follower output MOSFETs M9, M13, and M15. The sources of the source follower output MOSFETs M9, M13, and M15 are provided with the n-channel current source MOSFETs M10, M14, and M16 as loads. The above-mentioned three source follower output circuits form output signals that are electrically separated and the source output of the output MOSFET M9, which is one of those mentioned above, constitutes the feedback circuit of the amplification MOSFET M11 and is connected to the phase compensation capacitor C1.

The other two output MOSFETs are connected to the output terminals OUT1 and OUT2, respectively, and the output terminal OUT1 is used to output the output voltage so that the voltage of the drain of the neuron MOSFET and that of drain of the p-channel MOSFET are equal at the connection node as mentioned above, although not restricted particularly. The output terminal OUT2 is used to form the signal Vsi to be supplied to the C-matrix, which is the circuit in the next stage. Oscillation caused by the capacitance of the C-matrix in the next stage can be thus avoided.

The circuit diagram of the C-matrix of an embodiment is shown in FIG.8. The C-matrix circuit in the present embodiment has a structure in which capacitors are arranged in a matrix form and comparators are connected, and performs the operation to discriminate the sign of the results of the matrix operation as shown in expressions 15 and 16.

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{pmatrix} \quad \ldots (15)$$

$$z_t = \begin{cases} 1 & r_t > 0 \\ 0 & r_t < 0 \end{cases} \quad (t = 1, 2, \cdots, n) \quad \ldots (16)$$

$s = (s1, s2, \ldots, sm)^T$ is an m-dimensional input vector the components of which are positive, and zt is the component of the n-dimensional output vector $z = (z1, z2, \ldots, zn)^T$. The weighting matrix is an n x m matrix and the components wti can be negative or positive. The C-matrix has m comparison capacitors and the capacitance Ccmpi (i = 1, 2, ..., m) can be obtained by expressions 17 and 18.

$$C_{cmpi} = \begin{cases} C_0 & w_{mini} \geq 0 \\ C_0 - Cw_{mini} & w_{mini} < 0 \end{cases} \quad \ldots (17)$$

$$w_{mini} = \min \{w_{1i}, w_{2i}, \cdots, w_{ni}\} \quad \ldots (18)$$

According to the design rules, C_0 in expression 17 is the minimum capacitance and C is a step of available capacitance. When the difference between the minimum value wmini and the second minimum w in the same column is equal to or more than C_0/C, C_0 can be ignored and the comparison capacitor is determined simply by expression 19.

$$C_{cmpi} = \begin{cases} 0 & w_{mini} \geq 0 \\ -Cw_{mini} & w_{mini} < 0 \end{cases} \quad \ldots (19)$$

Other capacitors Cti (t = 1, 2, ..., n) (i = 1, 2, ..., m) are determined by expression 20 using the value Ccmpi of the comparison capacitor.

$$C_{ti} = Cw_{ti} + C_{cmpi} \quad \ldots (20)$$

In addition, dummy capacitors Cdumt (t = 1, 2, ..., n) are provided so that the summed value of the capacitors in each row is equal to the same value Csum.

The circuit diagrams to illustrate the operation method of the C-matrix circuit are shown in FIG.9A and

FIG.9B. In the operation method of the C-matrix circuit, all the MOSFET switches are turned on first, all the input voltages are set to 0 V, and the voltage of the floating node is pre-charged to 0 V, as shown in FIG.9A. Then, as shown in FIG.9B, all the MOSFETs are turned off to terminate the pre-charge and the input voltage Vini in proportion to each input component si is added. As a result, the potential of the comparison floating node is obtained as expression 21 and that of the t-th floating node is as expression 22.

$$V_{cmp} = \frac{\sum_{i=1}^{m} C_{cmpi} V_{ini}}{C_{sum}} \quad \ldots (21)$$

$$V_t = \frac{\sum_{i=1}^{m} C_{wti} V_{ini} + \sum_{i=1}^{m} C_{cmpi} V_{ini}}{C_{sum}} \quad \ldots (22)$$

If it is assumed that the output of the t-th comparator that compares these two potentials is Vdd, expression 23 is required because Vcmp < Vt, and it is found that this is the same operation as those shown by the above-mentioned expressions 15 and 16.

$$\sum_{i=0}^{m} C_{wti} V_{ini} > 0 \quad \ldots (23)$$

Since the speech recognition device relating to the present invention has the object to be applied to speech recognition, the spectrum envelopes of five vowels expressed in a feminine voicels are used as inputs to the present circuit. More concretely, the 30-dimensional vectors, each component of which is a rounded whole number from 1 to 255, are used. As a result of learning, the scale of this circuit is p = 30, m = 15, and n = 5 in the FIG.3. The circuit has been designed based on the values of the pattern vectors and weight vectors obtained from this learning.

In FIG.10, examples of the capacitance (fF) of the template value C1ij of the clustering layer when the five vowels (a, i, u, e, o) are recognized as mentioned above are shown. Capacitance C2ij is obtained by C2ij = 255 -

Clij. The node number corresponds to the 30-dimensional vector that corresponds to the above-mentioned spectrum envelope.

In FIG.11, examples of learned results of weight and the capacitance (fF) of C-matrix of the labeling layer when the five vowels (a, i, u, e, o) are recognized as mentioned above are shown.

The results of the simulation are shown in FIG.12, when the clustering layer and the labeling layer of a speech recognition device are configured in the above-mentioned structure and the five vowels (a, i, u, e, o) are entered. In the figure, the potentials of the comparison floating nodes that recognize /u/ of the C-matrix are shown. When a, i, u, e, and o are entered into the input in this order, the potential of the floating node is raised compared to the comparison com only for the input /u/ and a high level output signal Vout3 is output from the voltage comparison circuit.

In FIG.13, the output waveforms of the simulation result are shown, when the clustering layer and the labeling layer of the speech recognition device are configured as that in the above-mentioned structure and the five vowels (a, i, u, e, o) are entered. When a, i, u, e, and o are entered repeatedly in this order as input data, the output out "a", out "i", out "u", out "e", and out "o" are put out in this order. If the input data pointed by the arrow is assumed to be e, for example, the outputs out "a" to out "o" are put out as a digital signal with a pattern 0, 0, 0, 1, 0.

The speech recognition device relating to the present invention is designed with a clustering system of two inputs, four nodes and two outputs in accordance with the 1.5 μm rule. In order to digitize the input, the neuron MOSFET is made to have five inputs, and the ratio of the capacitances of four of them is designed to be 1: 2: 4: 8 to play a role of a simple digital/analog conversion. The chip area required for this design is

537,000μm^2.

In order to compare to the speech recognition device in the analog circuit structure relating to the present invention, designing with an 8-bit digital circuit is also carried out. In designing, the hardware description language Verilog-HDL is used. All operations are designed so as to be performed in parallel, similarly to the case of the analog circuit. The area required for this is 19,516,000μm^2. This indicates that the area can be reduced to one thirty-sixth, compared to that of the 8-bit digital circuit, if the above-mentioned analog circuit is used.

Although the larger the scale, the larger the chip area for wiring is required in a digital circuit, the larger the scale, the more advantage in area can be obtained in the speech recognition device of the present invention because of the structure in which the basic operation circuits are arranged in order.

Since the current/voltage characteristics of a MOSFET are used without modification in the speech recognition device relating to the present invention, a statistical analysis has been carried out in order to investigate how the variations in devices affect the cluster processing. The threshold voltages Vthn and Vthp of the n-channel MOSFET and the p-channel MOSFET are set based on a normal distribution with a standard deviation provided that σ= 0.1V, and the transconductance KPn and KPp, provided that σ= 10%, being independent parameters.

The amplifier circuit is designed using about 10 MOSFETs and it is assumed that these are arranged in a small area and the variations are small, then a set of Vthn, Vthp, KPn, and KPp is determined to be used as a typical value of the MOSFET in the amplifier circuit. Although capacitors are designed in accomplice with the limitations of the design rule that the minimum capacitance is 14 fF and the step is 1 fF, they are

varied at a ratio of σ= 1 fF regardless of capacitance. With these conditions, a set of data (a, i, u, e, o) is entered and a Monte-Carlo simulation is carried out 30 times. As a result, it is found that precise operations are ensured due to the redundancy of clustering even if there exist errors in the devices.

Although the present invention is described with reference to embodiments as above, it is obvious that the present invention is not restricted to the above-mentioned embodiments and various modifications are available without deviating from the concept. For example, it is an acceptable case in which the comparison capacitor is omitted in the C-matrix, a voltage follower circuit is provided at the output to put out the matrix operation outputs, and a level discriminating circuit to select the largest one among them is provided.

When consonants, voiced sounds, and semivoiced sounds are recognized in addition to the above-mentioned vowels, clustering layers using the above-mentioned neuron MOSFET or labeling layers using the C-matrix are provided in accordance with them. In this case, the multi-dimensional vector corresponding to the spectrum envelope of input is common to all the circuits and the input capacitance of the clustering layer becomes large. It is recommended, therefore, to divide the clustering layer into plural circuits and provide an input buffer circuit corresponding to each circuit. The present invention can be widely used as a speech recognition device composed of semiconductor integrated circuits.

The effects obtained from the typical examples of the present invention are briefly described below. The similarity circuits, which receive input signals composed of multi-dimensional vectors corresponding to the spectrum envelope of speech inputs to be recognized and put out characteristics based on the self-organizing algorithm, calculate a distance for a dimension using a pair of neuron MOSFETs corresponding to each dimension in

order to obtain the distance between the above-mentioned multi-dimensional input vectors and the pattern vectors prepared in advance for speech recognition, perform the clustering process by summing current that flows through each neuron MOSFET and forming a voltage signal that corresponds to the degree of similarity, supply the voltage signal to a matrix circuit for matrix operation in which capacitors corresponding to weighting operations are arranged in matrix, and perform the labeling process by putting out what is most similar to the patterns prepared in advance among the matrix operation outputs as the recognition result. Therefore, speech recognition can be realized in a small-scale circuit.